Cooperative Data Protection for Topology-Aware Decentralized Storage Networks

Siyi Yang[1], Ahmed Hareedy[2], Robert Calderbank[2], and Lara Dolecek[1]
[1] Electrical and Computer Engineering Department, University of California, Los Angeles, Los Angeles, CA 90095 USA
[2] Electrical and Computer Engineering Department, Duke University, Durham, NC 27708 USA
siyiyang@ucla.edu, ahmed.hareedy@duke.edu, robert.calderbank@duke.edu, and dolecek@ee.ucla.edu

***Abstract*—While codes with hierarchical locality have been intensively studied in the context of centralized cloud storage due to their effectiveness in reducing the average reading time, those in the context of decentralized storage networks (DSNs) have not yet been discussed. In this paper, we propose a joint coding scheme where each node receives extra protection through the cooperation with nodes in its neighborhood in a heterogeneous DSN with any given topology. Our proposed construction not only supports desirable properties such as scalability and flexibility, which are critical in dynamic networks, but also adapts to arbitrary topologies, a property that is essential in DSNs but has been overlooked in existing works.**

I. Introduction

Decentralized cloud storage networks (DSNs) such as Storj, in which no entity is solely responsible for all data, have emerged as a secure and economic alternative to centralized cloud services. DSNs are believed to be economically attractive since extra capacity can be afforded by utilizing idle storage space on devices at the edge of the network. Modern non-volatile memory technologies such as persistent memories have potential to boost the performance of next-generation DSNs due to their low latency in addition to their high endurance and reliability. While erasure-correction (EC) codes are widely used to combat component failures, EC schemes that address practical challenges such as varying topologies, high churn rates, heterogeneous bandwidths and link speeds, and dynamic node balancing for content delivery of hot files [3], are relatively overlooked. Existing coding solutions either consider homogeneous components [1], [2], or assume simple network topologies [3], [4].

In this paper, we propose EC solutions that are tailored to tackle those challenges pertaining to DSNs. In Section II, we introduce the DSN model and necessary preliminaries. In Section III, we define the notions of EC hierarchies and compatible cooperation graphs. In Section IV, we propose an explicit construction of hierarchical codes for nodes with their cooperation graph being compatible. Finally, we summarize our results and discuss future directions in Section V.

All the details can be found in [5] and [6].

II. System Model

As shown in Fig. 1, a DSN is modeled as a graph $G(V, E)$ and codewords are stored among the nodes in a cluster. A failed node in a cluster is regarded as an erased symbol in the codeword stored in this cluster. A cluster is represented by its master node solely. Each edge $e_{i,j} \in E$ represents a communication link connecting node v_i and node v_j, through which v_i and v_j are allowed to exchange information. Messages $\{\mathbf{m}_i\}_{v_i \in V}$ are jointly encoded as $\{\mathbf{c}_i\}_{v_i \in V}$, and $\mathbf{c}_i$ is stored in v_i. For a DSN denoted by $G(V, E)$, let $p = |V|$. Suppose G is associated with a tuple $(\mathbf{n}, \mathbf{k}, \mathbf{r}) \in (\mathbb{N}^p)^3$, where $\mathbf{k}, \mathbf{r} \succ \mathbf{0}$ and $\mathbf{n} = \mathbf{k} + \mathbf{r}$. Here, k_i represents the length of the message $\mathbf{m}_i$ associated with $v_i \in V$; n_i and r_i denote the length of $\mathbf{c}_i$ stored in v_i and its syndrome, respectively.

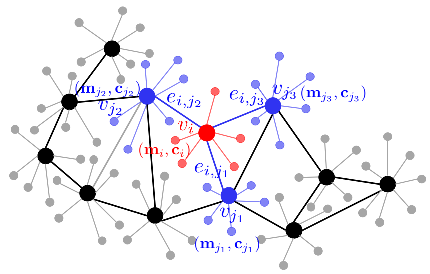

Fig. 1. Decentralized storage network (DSN). In node i, message $\mathbf{m}_i$ is encoded to $\mathbf{c}_i$.

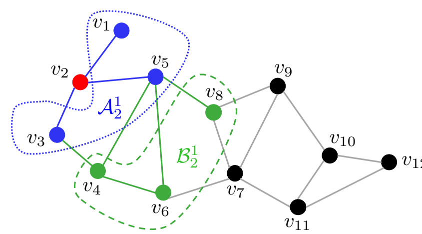

Fig. 2. DSN for Example 1.

III. Cooperative Data Protection

Denote the **EC hierarchy** of node $v_i \in V$ by a sequence $\mathbf{d}_i = (d_{i,0}, d_{i,1}, \dots, d_{i,L_i})$, where L_i is called the **depth** of $\mathbf{d}_i$, and $d_{i,l}$ represents the maximum number of erased symbols v_i can recover in its codeword $\mathbf{c}_i$ from the l-th level cooperation, for all $l \in [L_i]$. The 0-th level cooperation refers to local erasure correction, i.e., the local node v_i recovers its data without communicating with any neighboring nodes.

For each $v_i \in V$ such that $L_i > 0$, there exist two series of sets of nodes, denoted by $\varnothing \subset \mathcal{A}_i^1 \subset \mathcal{A}_i^2 \subset \cdots \subset \mathcal{A}_i^{L_i} \subseteq V$ and $\{\mathcal{B}_i^l\}_{l=1}^{L_i}$, where $\mathcal{A}_i^l \cap \mathcal{B}_i^l = \varnothing$ for all $l \in [L_i]$, and a sequence $(\lambda_{i,l;\mathcal{W}})_{\varnothing \subseteq \mathcal{W} \subseteq \mathcal{B}_i^l}$. In the l-th level cooperation, node $v_i \in V$ tolerates $\lambda_{i,l;\mathcal{W}}$ ($\varnothing \subseteq \mathcal{W} \subseteq \mathcal{B}_i^l$) erasures if all nodes in $\mathcal{A}_i^l \cup \mathcal{W}$ are able to decode their own messages, where the maximum value is $\lambda_{i,l;\mathcal{B}_i^l} = d_{i,l}$ and is reached when $\mathcal{W} = \mathcal{B}_i^l$; the minimum value is $\lambda_{i,l;\varnothing}$ and is reached when $\mathcal{W} = \varnothing$.

Example 1. *Consider the DSN shown in Fig. 2. For node* v_2, $\mathcal{A}_2^1 = \{v_1, v_3, v_5\}$, $\mathcal{B}_2^1 = \{v_4, v_6, v_8\}$. *Then,* v_2 *is able to correct* $d_{2,0}$, $\lambda_{2,1;\varnothing}$, $\lambda_{2,1;\mathcal{B}_2^1} = d_{2,1}$ *erasures by accessing nodes* v_2, *and* $\mathcal{A}_2^1$, $\mathcal{A}_2^1 \cup \mathcal{B}_2^1$ *in the first level cooperation, respectively.*

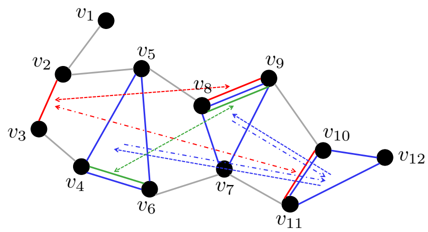

Fig. 3. Cooperation graph of Example 2.

Based on the aforementioned notation, for each $v_i \in V$ and $l \in [L_i]$, let $\mathcal{I}_i^l = \mathcal{A}_i^l \setminus \mathcal{A}_i^{l-1}$. The matrix $\mathbf{D} \in \mathbb{N}^{p \times p}$, in which $\mathbf{D}_{i,j}$ is equal to l for all $i, j \in [p]$ such that $j \in \mathcal{I}_i^l$, and zero otherwise, is called the **cooperation matrix**. A cooperation matrix is the adjacency matrix of a **cooperation graph**.

Example 2. *Fig. 3 is a cooperation graph corresponding to the cooperation matrix presented in the left panel in Fig. 4.*

Observe that the cooperation graph shown in Fig. 3 satisfies a set of conditions that define the so-called **compatible graph**.

Definition 1. *Let $\mathcal{G}$ be a cooperation graph on $G(V, E)$ that is represented by $\{(X_t, Y_t, \{X_{t;j}\}_{j \in Y_t}, \{Y_{t;i}\}_{i \in X_t}, l_t)\}_{t \in [T]}$, and let the 1-st level cooperation graph be $\mathcal{G}_1$. Let $\mathcal{M}_i$ denote the set of nodes that have an outgoing edge pointing at v_i in $\mathcal{G}_1$. For each $v_i \in V$, $l \in [L_i]$, let $R_{i;l} = \{t : i \in Y_t, t \in [T], l_t = l\}$, $R_i = \bigcup_{l \in [L_i]} R_{i;l}$, $T_{i;l} = \{t : j \in R_{i;l} \cap Y_{t;i}, t \in [T]\}$, and $V_{i;l} = \bigcup_{t \in R_{i;l}} Y_t$. We call $\mathcal{G}$ a **compatible graph** on G if the following conditions are satisfied:*

1) *For each $v_i \in V$, sets in $\{Y_t\}_{t \in R_i}$ are disjoint;*
2) *For each $v_i \in V$, $l \in [L_i]$, and any node v_j such that $j \in V_{i;l} \setminus \{v_i\}$, $V_{j;l} \subseteq \mathcal{M}_i$.*

IV. Code Construction

In cooperative data protection, we let $\mathbf{c}_i = [\mathbf{m}_i | \mathbf{s}_i]$, where $\mathbf{s}_j = \sum_{i \in [p] \setminus \{j\}} \mathbf{m}_j \mathbf{A}_{i,j}$, for some $\{\mathbf{A}_{i,j} \in \mathrm{GF}_q^{k_i \times r_i}\}_{i,j=1}^{p}$, is the parity of $\mathbf{m}_i$, $i \in [p]$. We show in Construction 1 and Theorem 1 the existence of a hierarchical coding scheme with cooperation graph $\mathcal{G}$ if $\mathcal{G}$ is a compatible graph.

Construction 1. *Given a DSN $G(V, E)$ with parameters $(\mathbf{n}, \mathbf{k}, \mathbf{r})$. Suppose $\mathcal{G}$ is a compatible graph of depth L on G, with parameters $\{(X_t, Y_t, \{X_{t;j}\}_{j \in Y_t}, \{Y_{t;i}\}_{i \in X_t}, l_t)\}_{t \in [T]}$, and the 1-st level cooperation graph is denoted by $\mathcal{G}_1$.*

Let $\boldsymbol{\delta}$ be the 1-st level cooperation parameter. For each $v_i \in V$, $1 \le l \le L_i$, and any $t \in T_{i;l}$, assign cooperation parameter $\gamma_{i;t}$ to C_t; let $\eta_{j;l} = \max_{t \in R_{j;l}, i \in X_{t;j}} \gamma_{i;t}$. Let $u_i = k_i + \delta_i + \sum_{l=2}^{L_i} \eta_{i;l}$, $v_i = r_i + \sum_{j \in \mathcal{M}_i} \delta_j + \sum_{2 \le l \le L_i, t \in T_{i;l}} \gamma_{i;l;t}$, for $i \in [p]$. For each $i \in [p]$, let $a_{i,s}, b_{i,t}$, $s \in [u_i]$ and $t \in [v_i]$ be distinct elements of $\mathrm{GF}(q)$, where $q \ge \max_{i \in [p]} \{u_i + v_i\}$.

Consider the Cauchy matrix $\mathbf{T}_i$ on $\mathrm{GF}(q)^{u_i \times v_i}$ such that $\mathbf{T}_i = \mathbf{Y}(a_{i,1}, \ldots, a_{i,u_i}; b_{i,1}, \ldots, b_{i,v_i})$, for $i \in [p]$. Then, we obtain $\mathbf{A}_{i,i}$, $\mathbf{B}_{i,j}$, $\mathbf{E}_{i;l}$, $\mathbf{U}_i$, $\mathbf{V}_{i;l}$, for $i \in [p]$, $j \in [p] \setminus \{i\}$, according to the following partition of $\mathbf{T}_i$,

$$\mathbf{T}_i = \left[\begin{array}{c|c|c|c|c} \mathbf{A}_{i,i} & \mathbf{B}_i & \mathbf{E}_{i;2} & \ldots & \mathbf{E}_{i;L_i} \\ \hline \mathbf{U}_i & & & & \\ \hline \mathbf{V}_{i;2} & & & & \\ \hline \vdots & & \mathbf{Z}_i & & \\ \hline \mathbf{V}_{i;L_i} & & & & \end{array} \right], \qquad (5)$$

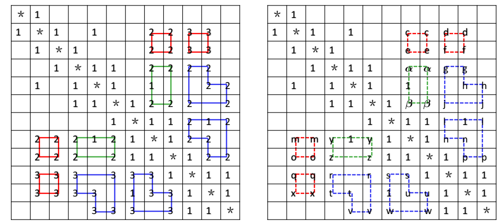

Fig. 4. Matrices $\mathbf{D}$ (left) and $\mathbf{X}$ (right).

$$\textit{where} \quad \mathbf{B}_i = \left[\begin{array}{c|c|c} \mathbf{B}_{i,j_1} & \ldots & \mathbf{B}_{i,j_{|\mathcal{M}_i|}} \end{array} \right], \qquad (6)$$

$$\textit{and} \quad \mathbf{E}_{i;l} = \left[\begin{array}{c|c|c} \mathbf{E}_{i;l;t_1} & \ldots & \mathbf{E}_{i;l;t_{|T_{i;l}|}} \end{array} \right], \qquad (7)$$

such that $\mathcal{M}_i = \{j_1, j_2, \ldots, j_{|\mathcal{M}_i|}\}$, $T_{i;l} = \{t_1, t_2, \ldots, t_{|T_{i;l}|}\}$, $\mathbf{A}_{i,i} \in \mathrm{GF}(q)^{k_i \times r_i}$, $\mathbf{U}_i \in \mathrm{GF}(q)^{\delta_i \times r_i}$, $\mathbf{V}_{i;l} \in \mathrm{GF}(q)^{\eta_{i;l} \times r_i}$, $\mathbf{B}_{i,j} \in \mathrm{GF}(q)^{k_i \times \delta_j}$ for all $v_j \in \mathcal{M}_i^1$, and $\mathbf{E}_{i;l;t} \in \mathrm{GF}(q)^{k_i \times \gamma_{i;t}}$. Let $\mathbf{B}_{i,j} = \left[\mathbf{E}_{i;l;t}, \mathbf{0}_{k_i \times (\eta_{j;l} - \gamma_{i;t})} \right]$, and $\mathbf{A}_{i,j} = \mathbf{B}_{i,j} \mathbf{V}_{j;l}$, for all $j \in Y_{t;i}$, $t \in T_{i;l}$; let $\mathbf{A}_{i,j} = \mathbf{B}_{i,j} \mathbf{U}_j$ for $v_j \in \mathcal{M}_i$; otherwise $\mathbf{A}_{i,j} = \mathbf{0}_{k_i \times r_i}$. The resulting code is $\mathcal{C}_2$, and its generator matrix $\mathbf{G}$ is constructed by substituting in [5, (1)].

Theorem 1. *The code $\mathcal{C}_2$ has EC hierarchies $\mathbf{d}_i = (d_{i,0}, d_{i,1}, \ldots, d_{i,L_i})$, for all $v_i \in V$, where $d_{i,0} = r_i - \delta_i - \sum_{l=2}^{L_i} \eta_{i;l}$, $d_{i,1} = r_i + \sum_{j \in \mathcal{M}_i} \delta_j$, and $d_{i,l} = r_i + \sum_{j \in \mathcal{M}_i} \delta_j + \sum_{2 \le l' \le l, t \in T_{i;l'}} \gamma_{i;t}$. Moreover, $\mathcal{I}_i^1 = \mathcal{M}_i$, $\mathcal{B}_i^1 = \bigcup_{v_j \in \mathcal{M}_i} (\mathcal{M}_j \setminus (\{v_i\} \cup \mathcal{M}_i))$; for $2 \le l \le L_i$, $\mathcal{I}_i^l = \bigcup_{t \in R_{i;l}} X_{t;i}$, $\mathcal{B}_i^l = \bigcup_{v_j \in \mathcal{I}_i^l} \left(\mathcal{I}_j^l \setminus (\{v_i\} \cup \mathcal{A}_i^l) \right)$ ($\mathcal{A}_i^l = \bigcup_{l' \le l} \mathcal{I}_i^{l'}$), $\lambda_{i,l;\mathcal{W}} = r_i + \sum_{v_j \in \mathcal{M}_i, (\mathcal{M}_j \setminus \{v_i\}) \subseteq (\mathcal{M}_i \cup \mathcal{W})} \delta_j + \sum_{\substack{2 < l' \le l, t \in T_{i;l}, C_t = l'', Y_{t;i} = \{j, j'\}, \\ \mathcal{I}_j^{l''} \setminus \mathcal{A}_i^{l'} \subseteq (\{v_i\} \cup \mathcal{W}) \text{ or } \mathcal{I}_{j'}^{l''} \setminus \mathcal{A}_i^{l'} \subseteq (\{v_i\} \cup \mathcal{W})}} \gamma_{i;t}$, for $\emptyset \subseteq \mathcal{W} \subseteq \mathcal{B}_i^l$.*

Example 3. *In Fig. 4, label nodes on each horizontal edge in each cycle in $\mathbf{D}$ by a unique symbol s to obtain $\mathbf{X}$. Obtain $\{\mathbf{U}_i\}_{i \in [12]}$, $\{\mathbf{V}_{i;l}\}_{i \in [12], l \in \{2,3\}}$, $\mathbf{B}_s$, according to Construction 1. Then, $\mathbf{A}_{i,j} = \mathbf{V}_{i;l} \mathbf{B}_s$, for each (i, j) that has $\mathbf{X}_{i,j} \neq 0$.*

V. Conclusion and Future Work

Hierarchical locally accessible codes in the context of centralized networks have been discussed in various prior works, whereas those of DSNs have not been explored. In this paper, we presented a cooperative data protection scheme for DSNs, which adapts to arbitrary network topologies. Our scheme supports important properties in cloud storage systems, which are scalability, heterogeneity, and flexibility.

References

[1] H. Hou, P. P. C. Lee, K. W. Shum, and Y. Hu, "Rack-aware regenerating codes for data centers," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 4730–4745, Aug. 2019.

[2] N. Prakash, V. Abdrashitov, and M. Médard, "The storage versus repair-bandwidth trade-off for clustered storage systems," *IEEE Transactions on Information Theory*, vol. 64, no. 8, pp. 5783–5805, Aug. 2018.

[3] J. Li, S. Yang, X. Wang, and B. Li, "Tree-structured data regeneration in distributed storage systems with regenerating codes," in *Proceedings of 2010 IEEE INFOCOM*, 2010, pp. 1–9.

[4] J. Pernas, C. Yuen, B. Gastón, and J. Pujol, "Non-homogeneous two-rack model for distributed storage systems," in *IEEE International Symposium on Information Theory*, Jun. 2013, pp. 1237–1241.

[5] S. Yang, A. Hareedy, R. Calderbank, and L. Dolecek, "Topology-aware cooperative data protection in blockchain-based decentralized storage networks," in *2020 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2020, pp. 622–627.

[6] ——, "Hierarchical coding for cloud storage: Topology-adaptivity, scalability, and flexibility," 2020. [Online]. Available: https://arxiv.org/abs/2009.09146